

SONDAGES EN LIGNE (2 / 3)

Des problèmes méthodologiques importants pour les sondages en ligne

1. En ligne, un échantillon aléatoire est difficile – voire impossible – à constituer. Il est impossible de garantir à chaque membre de la population ciblée une probabilité égale de faire partie de l'échantillon dans les cas, par exemple, où l'on veut sonder un groupe d'âge (les adultes) ou un territoire géographique. On sait que tous les individus ciblés n'ont pas accès à Internet ou n'utilisent pas Internet. À titre indicatif, et selon la province ciblée, le taux d'utilisation régulière d'Internet¹ par les adultes se situe entre 70 % et 80 %. Pour la province de Québec - selon le Cefrio – il était de 74,4 % au Québec pour décembre 2010; le maximum avait été atteint en juillet 2010 avec 77 %. Environ 25 % des adultes québécois auraient ainsi une probabilité quasi nulle ou nulle de faire partie de l'échantillon d'un sondage les concernant.

2. Selon l'approche utilisée, il est difficile ou impossible de contrôler l'échantillon, ce qui résulte encore une fois en un échantillon qui ne garantit pas à tous les membres de la population ciblée une chance égale d'en faire partie. Voici quelques cas particuliers.

- Un même répondant peut répondre plusieurs fois en contournant les obstacles technologiques (souvent faciles à contourner par ailleurs), en répondant à la place de quelqu'un d'autre ou en répondant alors qu'il ne le devrait pas (ne fait pas partie de la population ciblée). Pensons par exemple à un sondage qui s'intéresse à l'opinion de résidents d'une municipalité, mais qui valide la participation en utilisant l'adresse Internet des répondants. Il serait alors impossible de contrôler l'échantillon. Un même répondant pourrait participer plusieurs fois en se créant à répétition des adresses de courriel (Hotmail, Gmail, Yahoo, etc.), en supprimant les cookies de son ordinateur et en réinitialisant sa mémoire cache, en changeant d'ordinateur pour répondre, etc.
- La liste de contacts peut être incomplète, avec des doublons, avec des adresses de courriels non valides, en pointant vers des personnes qui ne font pas partie de la population ciblée ou avec plus d'une adresse différente pointant vers une même personne, donnant à celle-ci la possibilité de répondre au sondage plus d'une fois.
- Un sous groupe de la population peut faire l'objet d'incitatifs à répondre qui font en sorte qu'il sera sur-représenté (mobilisation partisane par exemple).
- Un sous groupe de la population peut ignorer qu'un sondage auquel il pourrait participer a lieu.

Autres problèmes potentiels :

- la validation de l'identité du répondant (se qualifie-t-il ?) ;
- le maintien du niveau de sérieux des répondants ;
- les invitations à participer à un sondage qui sont traitées comme du spam par des serveurs.

En somme, on en revient toujours à un échantillon qui n'aura pas été constitué en accordant à tous les membres de la population ciblée une probabilité égale et non nulle d'être sélectionnés. Les principales conséquences sont que les résultats peuvent être biaisés (ce sont les résultats qui sont biaisés et non le questionnaire ou le sondage), mais surtout que dans les cas où l'on doit conclure qu'un échantillon est non aléatoire, l'inférence statistique est exclue, pour au moins deux raisons : premièrement, il n'est pas possible de calculer une marge d'erreur pour un échantillon non aléatoire et deuxièmement, puisque l'échantillon n'est pas aléatoire, il est impossible de prétendre – statistiquement – que ses caractéristiques correspondent à celles de la population dont il provient.

Note

Un sondage dont les répondants n'ont pas été sélectionnés de façon aléatoire NE DOIT PAS afficher de marge d'erreur. Il s'agit d'une norme importante de l'ARIM². Dans la plupart des sondages en ligne (et aussi des sondages par *intercept* – personne à personne, dans la rue par exemple) les répondants ne sont pas sélectionnés aléatoirement et la notion de marge d'erreur devient donc caduque. Pour une discussion sur les échantillons aléatoires (probabilistes) et non aléatoires (non probabilistes), voir Info Takto de juillet / août 2010.

¹ Utilisation d'Internet au moins une fois par semaine pendant la période évaluée.

² Association de la recherche et de l'intelligence marketing : www.mria-arim.ca